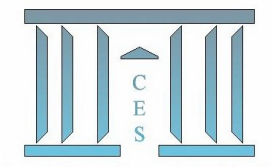
Dumbing down rational players: learning and teaching in an experimental game

Antoine Terracol, Jonathan Vaksmann

Documents de Travail du
Centre d'Economie de la Sorbonne

CES Working Papers



Dumbing down rational players : Learning and teaching in an experimental game

Antoine TERRACOL, Jonathan VAKSMANN

2007.17

CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE

Maison des Sciences Économiques, 106-112 boulevard de L'Hôpital, 75647 Paris Cedex 13
http://ces.univ-paris1.fr/CESPublicat.htm
ISSN en cours d'attribution

Dumbing down rational players: Learning and teaching in an experimental game

Antoine Terracol* and Jonathan Vaksmann†

March 28, 2007

Abstract

This paper uses experimental data to examine the existence of a teaching strategy among bounded rational players. If players realize that their own actions modify their opponents' beliefs and actions, they might play certain actions to this specific end; and forego immediate payoffs if the expected payoff gain from a teaching strategy is high enough. Our results support the existence of a teaching strategy in several ways: First they show that players update their beliefs in order to take account of the reaction of their opponents to their own action. Second, we examine if players actually use a teaching strategy by playing an action that induces a poor immediate payoff but is likely to modify the opponent's behavior so that a preferable outcome might emerge in the future. We find strong evidence of such a strategy in the data and confirm this finding within a logistic model which suggests that the future expected payoff that could arise from a teaching strategy has indeed a significant impact on choice probabilities. Finally, we investigate the effective impact of a teaching strategy on achieved outcomes and find that efficient teachers can successfully use teaching in order to reach their favorite outcome at the expense of their opponents.
JEL codes: C72, C91, D83
Keywords: Game theory, Teaching, Beliefs, Experiment

Résumé

Ce papier utilise des données expérimentales pour analyser l'existence d'une stratégie de *teaching* de la part de joueurs à rationalité limitée. Si les joueurs réalisent que leurs propres actions modifient les croyances et les actions de leurs opposants, ils peuvent être amenés à choisir certaines actions dans le but de modifier le comportement de leurs opposants dans le futur quand bien même ces actions ne maximiseraient pas leurs paiements immédiats. Nos résultats supportent

*GREMARS (Université Lille 3); and Université Paris 1 Panthéon-Sorbonne, CNRS. Centre d'Économie de la Sorbonne, 106-112 boulevard de l'Hôpital, 75642 Paris cedex 13, France, terracol@univ-paris1.fr

†Université Paris 1 Panthéon-Sorbonne, CNRS. Centre d'Économie de la Sorbonne, 106-112 boulevard de l'Hôpital, 75642 Paris cedex 13, France, jonathan.vaksmann@univ-paris1.fr
We wish to thank Jordi Brandts, Maxim Frolov, Claude Meidinger, Bruno Rock and participants in various seminar for their help and valuable discussions, and the Centre National de la Recherche Scientifique for financial support. All remaining errors are ours.

l'existence d'une stratégie de teaching de plusieurs manières : tout d'abord, ils montrent que les joueurs actualisent leurs croyances afin de tenir compte de la réaction de leurs opposants à leurs propres actions. Deuxièmement, nous examinons si les joueurs utilisent effectivement une stratégie de teaching en choisissant des actions associées à des paiements immédiats relativement faibles mais susceptibles de modifier le comportement futur de leurs opposants. Nous trouvons que ce type de stratégie est cohérent avec nos données et nous confirmons ce résultat à l'aide d'un modèle logistique qui indique que le gain futur attendu associé à une stratégie de teaching a en effet un impact significatif sur les probabilités de choix des joueurs. Enfin, nous étudions les conséquences de l'utilisation d'une stratégie de teaching sur les issues atteintes dans le jeu et nous trouvons que les teachers efficaces peuvent avec succès utiliser une stratégie de teaching afin d'atteindre l'issue qui leurs procure le gain le plus élevé aux dépens de leurs opposants.
Mots-clés: Théorie des jeux, Teaching, Croyances, Expérimentation

1 Introduction and motivations

Traditional game theory focuses on strategic interactions among fully rational players. But, as is now widely recognized, humans' reasoning might be bounded. Thus, many approaches have suggested ways to encompass these limits into game theory. While some studies address the theoretical side of the question by examining, for example, convergence supported by evolutionary forces or adaptive rules; other approaches, based on empirical standards, typically aim to use experimental data to provide a more accurate description of players' behavior. In the present study we aim to contribute to the latter framework.

In the part of the literature designed to describe the way people play their game, several studies analyze the interactions among adaptive players who choose their response according to what they have experienced in the past. Some studies (e.g. Roth and Erev, 1995; and Arthur, 1991) focus on reinforcement learning where players determine their actions according to their success in bringing high payoffs in the previous periods of the game. In belief-based learning models (Cheung and Friedman, 1997; Boylan and El-Gamal, 1993; Mookherjee and Sopher, 1994, 1997; Rankin et al., 2000; and Fudenberg and Levine, 1998) players use the past history of the game to update their beliefs about what their opponents may play. Finally, other studies (e.g. Camerer and Ho, 1999) take both reinforcement and belief learning as two components of the same learning model, known as the EWA learning model, which has been proven to be very successful in predicting how people behave. All these models provide foundations for equilibrium theory and offer opportunities to model empirically observed behavior.

Yet, one might think that these approaches are at odds with the foundations of traditional game theory by ruling out perfect rationality, and regarding people as merely adaptive. In this paper we attempt to empirically exhibit a behavior that stands between these two extremes.

The starting point of our study is that in the learning models, players' decision process has only a backward component: they respond according to what they have learned from the past. This rules out any awareness that opponents might also learn from other players' actions; and might thus be influenced by them. However, if a player suspects

that his opponent can also learn from the past history of the game, he might attempt to play strategically to drive him into a particular set of actions. That is how a *teaching strategy* (i.e. playing an action that does not necessarily maximize the expected payoff at the current round, but increases the probability of convergence to an equilibrium deemed preferable) might arise. Thus we aim to exhibit, among adaptive players, a piece of sophistication or, from another point of view, we are actually dumbing down rational players.

Very few studies have focused on the issue of teaching. A first interesting exception is Ehrblatt et al. (2006). The authors analyze how teaching speeds up convergence to a unique (pure strategy) equilibrium. They suggest that if a player's actions converge before his beliefs, he must have chosen the Nash action (despite the fact that this action is outside his best-response set) in order to teach his opponent. They then use a criterion to separate teachers and learners in pairs of players: the player whose actions converge first in a pair is said to be the teacher and the player whose actions converge later is a learner. In the present study, our approach is different: Rather than examining the speed of convergence to a unique equilibrium, we want to investigate teaching when both players have the ability to teach and are willing to lead their opponent to their favorite outcome. In order to make our purpose not trivial, we wanted players' favorite outcomes to be different and thus used a game with diverging interests.

A second exception is Camerer et al. (2002) who adds sophistication to adaptive models by assuming a heterogenous population of players: A fraction of them are supposed to be fully rational and consequently have the ability to exhibit equilibrium behavior. Those players have in mind their own estimation of the repartition (not necessarily the real one) between rational and adaptive players, and use their knowledge to outguess their adaptive opponents. The remaining fraction is regarded as adaptive and only looks backward to choose the current action. Hence the authors examine how people learn in a heterogenous population of players. In this paper, we adopt another approach by considering the strategic interaction in pairs of players who are both neither extremely sophisticated nor extremely unsophisticated.

In order to investigate the existence of such a forward-looking behavior, we use data from an experimental game with multiple equilibria in which players have diverging interests.

After describing our experimental design, we test whether players see their opponents as learners who observe the history of the game and modify their behavior accordingly; which indeed represents a necessary condition for players to adopt a teaching strategy. To this end, we estimate the difference between players' actual (or "true") beliefs, elicited using an appropriate scoring rule, and γ-weighted beliefs (Cheung and Friedman, 1997) where players only consider the past history of their opponents' actions. We then examine whether this difference can be explained by players' own past actions, i.e. we test whether they perceive that their own actions can modify their opponents' behavior.

Teaching implies that players could choose actions with poor immediate expected payoff but which are likely to influence the opponent's behavior so that they might lead to a more profitable outcome in the future. Thus, in a following step, we examine this implication in two parts. First, we test the existence of such strategy by examining whether players statistically depart from best responding by playing actions which

might lead to the emergence of a more favorable equilibrium in future. Second, we formally test if players choose their actions according to the prospective payoff gain that might follow from a modification of their opponents' beliefs. More precisely, we estimate a logistic model where the probability of a given action being played depends not only on the immediate expected payoff, as it is usually the case, but also on the cumulative future payoff gain induced by the expected modification of the opponent's behavior.

Finally, we examine the actual consequences of teaching and analyze whether using a teaching strategy is an effective tool for players to drive the outcome of a game.

2 Experimental design and procedures

The experiment was run using the computerized experimental laboratory of the University of Paris 1 Panthéon-Sorbonne from the Summer through the Fall of 2006.[1] No subject had any training in game theory. Each experimental session lasted almost one hour and a half. All sessions consists of 30 repetitions of the game represented below under two strategic treatments we will describe in this section. During the experiment, players were evenly divided into type-1 and type-2 players. Payoffs were denominated in units of experimental currency and converted into Euros at the end of each session. The subjects, on average, earned approximately €14.4 for their participation. They were paid €3 just for showing up.[2]

The game we used can be represented by the matrix below where row player is of type 1, while column player is of type 2. It is in fact a reduced form of the duopoly game initiated by Hamilton and Slutsky (1990). In their set-up, X,Y and Z refer respectively to the Cournot, the Stackelberg leader and the follower quantities.

Payoff matrix

		type 2		
		X	Y	Z
type 1	X	(40,52)	(22,46)	(40,52)
	Y	(35,40)	(10,20)	(44,46)
	Z	(40,52)	(30,60)	(40,52)

This game has many features we desired in our design. First of all, it is easy to understand. Then, it has three non Pareto-rankable Nash equilibria: (X,X), (Y,Z) and (Z,Y) which are not too difficult to calculate or learn deductively.

The multiplicity and non Pareto-rankability of these equilibria are useful features for our purpose. Indeed, as already noted, we are interested in investigating teaching incentives in a set-up where both players are potential teachers and are willing to lead their opponent to their favorite outcome. On the other hand, diverging interests make teaching more useful and consequently more likely to emerge because players have

[1] For conducting the experiment we used the experimental software 'Regate' (Zeiliger, 2000).

[2] Here, all payoffs will also be denominated in units of experimental currency. Subjects have been paid according to the sum of the payoffs they received during the 30 repetitions of the stage game. Every 200 units of experimental currency could be redeemed for €1.6 after each session. For more details, instructions for the subjects are available from the authors upon request.

conflicting preferences over equilibria. For instance, at (Y,Z), type-1 players get their best payoff. Thus, it is natural that type-1 players would like to make (Y,Z) emerge. But one can easily notice that at this equilibrium, the type-2 opponents get their worst equilibrium payoff, hence this equilibrium exhibits a strong conflict of interests. Likewise, (Z,Y) is the best outcome for type-2 players while it is the worst equilibrium in terms of payoff for the type-1 opponents. Then, consistent to the terminology used by Hamilton and Slutsky[3], if we observe convergence to (Y,Z) (resp. (Z,Y)) in a pair, one could say that type-1 (resp. type-2) player takes the leadership while the opponent is the follower. Consequently, throughout the paper, we will refer to (Y,Z) and (Z,Y) respectively as type-1 and type-2 leadership equilibria.

Finally, one last interesting feature is the use of asymmetric payoffs which generates interesting testable implications concerning relative teaching incentives across players.

For each session we conducted, we adopted one of two treatments that differ according to the matching protocol that was implemented. More precisely, we performed a fixed-opponent, or 'Partner', treatment and a random opponent, or 'Stranger', treatment. In both treatments, the type assigned to each player remains the same throughout the experiment. If players attempt to 'teach' their opponent, the way in which they are matched when a game is played repeatedly might affect behavior. Thus, running these two treatments allow us to test the impact of random matching. We recruited 76 subjects, 40 for the Partner treatment and 36 (two sessions of 18 subjects each) for the Stranger treatment.

In each round, before choosing their action, subjects' beliefs are elicited using a proper scoring rule defined below. They were asked to report their beliefs or prediction about the likelihood that their opponent would use each of his actions available in the current round, i.e. X, Y or Z.

We now describe more precisely the belief elicitation procedure. This procedure takes the classical quadratic form used in the literature. Subjects were asked to report the probability that their opponent will play X, Y or Z. Such a report takes the form of a vector $b = (b_X, b_Y, b_Z)$ where b_a represents the belief held by the subject associated to the action a of his opponent (a =X,Y,Z). A player's payoff when he reports b and when his opponent actually uses action a is given by $\left[8 - 4\left((1-b_a)^2 + \sum_{z\neq a} b_z^2\right)\right]$.

This formula means that each subject receives an endowment of 8 units of experimental currency at the beginning of each round and reports his beliefs. The amount $(1-b_a)^2$ subtracted from the initial endowment of the subject corresponds to a penalty for having reported an inappropriate belief for the action a his opponent has played in the current round. Note that this penalty equals 0 when the subject reports a probability $b_a = 1$ and his opponent plays a at the current round. Subjects are also penalized for having stated inappropriate beliefs for the other actions by a substraction of an amount $\sum_{z\neq a} b_z^2$ from their initial endowment of 8 units of experimental currency. The worst possible guess, i.e putting all the probability weight on an action that the opponent does not actually choose, leads a payoff of 0 (and explains the normalization constant (4) which appears in the formula). It can be easily demonstrated than according to this

[3]More precisely, in Hamilton and Slutsky, (X,X) corresponds to the Cournot equilibrium, while (Y,Z) and (Z,Y) correspond to the Stackelberg equilibria.

computation, risk neutral subjects should tell the truth. [4]

As usual in this kind of design, the reward for reporting beliefs remains small in comparison with the payoffs associated to the game. Indeed, this is an important point because a too large reward could affect subjects' behavior due to the possibility of playing any particular action repeatedly so as to maximize their prediction payoffs at the expense of their game payoffs.

3 Results

3.1 Player's beliefs

In order to use a teaching strategy, players must first be aware of the learning process of their opponents, i.e. they must be aware of the fact that their opponents use the past history of the game to form their beliefs. In this section, we test if players anticipate their opponents' reaction to their own action or, in other words, if they see their opponents as learners.

We assume that player i's "true" beliefs about what player j will play in round t, is determined by the sequence of actions played by his opponent from round 1 to round $t-1$, which we denote by $\{a_j(\tau)\}_{\tau=1}^{t-1}$; and is also determined by his own action in round $t-1$, $a_i(t-1)$ if he believes his opponent is a learner. Player i's true beliefs about the action a, a =X,Y,Z of player j in round t can thus be written as:

$$B_i^a\left(t|\{a_j(\tau)\}_{\tau=1}^{t-1}, a_i(t-1)\right)$$

Explicitly modeling the way a player's actions influence his own beliefs *via* his opponent's anticipated reaction would undoubtedly lead to an intricate model, and would require strong behavioral assumptions. Our aim in this section is not to describe accurately those complex interactions, but rather to test if players think of their opponents as learners who observe others' actions and modify their behavior accordingly. Our strategy will be to test if true beliefs $B_i^a(t)$ significantly differ from beliefs that would only depend on $\{a_j(\tau)\}_{\tau=1}^{t-1}$, the history of his opponent's past actions up to round $t-1$, and if this difference can be explained by player i's own action in the previous round.[5]If so, we could say that player i perceives that his own actions can modify his opponent's behavior.

Models of belief-learning in the literature typically assume that individuals form their beliefs on the basis of the history of their opponents' past actions[6], so that the "empirical" belief held by player i about player j playing action a in round t can be written as:

$$\tilde{B}_i^a\left(t|\{a_j(\tau)\}_{\tau=1}^{t-1}\right)$$

[4]Several studies (e.g. Sonnemans and Offerman, 2001, Nyarko and Schotter, 2002) indicate that, with this quadratic scoring rule, players indeed report the truth. Rutström and Wilcox (2006), however, find that an intrusive scoring rule for belief elicitation affects people's behavior.

[5]Note that the impact of less recent actions on the opponent's behavior is already encompassed in $\{a_j(\tau)\}_{\tau=1}^{t-1}$.

[6]A typical example would be the Bayesian updating rule.

Since $\tilde{B}_i^a(t)$ is conditional on $\{a_j(\tau)\}_{\tau=1}^{t-1}$, but not on $a_i(t-1)$, a test of wether i's true beliefs indeed depend on player i's actions would be to test if the difference between the true and "empirical" beliefs can be explained by $a_i(t-1)$. Formally, if true beliefs $B_i^a(t)$ truly depend on the player's past actions, then $B_i^a(t) - \tilde{B}_i^a(t)$ should depend on $a_i(t-1)$:

$$B_i^a\left(t|\{a_j(\tau)\}_{\tau=1}^{t-1}, a_i(t-1)\right) - \tilde{B}_i^a\left(t|\{a_j(\tau)\}_{\tau=1}^{t-1}\right) = R_i^a\left(t|a_i(t-1)\right)$$

Thus, if $B_i^a(t) - \tilde{B}_i^a(t)$ depends on $a_i(t-1)$, then true beliefs $B_i^a(t)$ must also depend on $a_i(t-1)$, and we may conclude that players think that their opponents modify their behavior according to the history of the game, and take this into account into their own beliefs.

We chose to specify $\tilde{B}_i^a(t)$ as γ-weighted "empirical" beliefs (Cheung and Friedman, 1997) where the belief held by player i about the probability that player j will play action a in round $t+1$ is given by:

$$\tilde{B}_i^a(t+1) = \frac{\mathbb{1}_{\{a_j(t)=a\}} + \sum_{u=1}^{t-1} \gamma^u \mathbb{1}_{\{a_j(t-u)=a\}}}{1 + \sum_{u=1}^{t-1} \gamma^u} \tag{1}$$

where $\mathbb{1}_{\{a_j(t)=a\}}$ equals one if player j has played action a in round t, and zero otherwise. Actions played in a given round are discounted with time at rate $\gamma \in [0,1]$. This model encompasses the Cournot model (for $\gamma = 0$) where the belief held in period t about action a is one if the strategy has been play in round $t-1$, and zero otherwise; and the fictitious play model (for $\gamma = 1$) where the belief about a given action corresponds to the frequency with which this action has been played since round 1. The Cheung and Friedman model has been found to perform well empirically to explain people's behavior in games.

As can be seen from Equation (1), γ-weighted beliefs in round t are only conditional on $\{a_j(\tau)\}_{\tau=1}^{t-1}$, the past history of the actions played by *other* players up to round $t-1$, and are thus good candidates for constructing $\tilde{B}_i^a(t)$.

To do so, we estimate the model of equation (1) at the individual level using the method of minimum mean-squared error[7] along the lines of Nyarko and Schotter (2002). We are thus able to compute estimated empirical beliefs $\hat{\tilde{B}}_i^a(t)$ that can be interpreted as the largest part of the individual's true beliefs $B_i^a(t)$ that can be explained by the past history of the game up to round $t-1$ under the Cheung-Friedman hypothesis. We then compute $\hat{R}_i^a(t)$, the difference between true (or elicited) and estimated empirical beliefs, and proceed to test whether these differences vary according to the action taken by the individual in the previous round. For each variable $\hat{R}_i^a(t)$, we test if its distribution differs according to whether i has played in $t-1$ the action to which a is a best response or not (X for $\hat{R}_i^X(t)$, Z for $\hat{R}_i^Y(t)$, and Y for $\hat{R}_i^Z(t)$). The first row of Table 1 reports the differences in the mean estimated residual $\hat{R}_i^a(t)$ according to whether the action for which a is a best response has been played in the previous round, i.e. it is the difference $\hat{D}_i^a(t) = \hat{R}_i^a(t|a_i(t-1) = \tilde{a}) - \hat{R}_i^a(t|a_i(t-1) \neq \tilde{a})$ where a is a best response to $\tilde{a}$. A positive $\hat{D}_i^a(t)$ means that players' true beliefs regarding a

[7] That is, our estimator is the values of the parameter vector that minimize $\sum_{i,t,a}\left(B_i^a(t) - \tilde{B}_i^a(t)\right)^2$.

given action have a stronger upward bias (or a weaker downward bias) when the action for which it is a best response has been played in the previous round.

The following 3 rows present various test statistics (t-test with unequal variance, Mann-Whitney, and Somer's D statistic) for the equality of means of $\hat{R}_i^a(t)$ in the Partner and Stranger treatments (the corresponding p-values are given in parentheses below the test statistic).

The $\bar{\hat{D}}_i^a(t)$ are always larger in the Partner than in the Stranger treatment, which is consistent with the fact that the former allows players to influence their opponents' choices by playing specific actions. The fact that the means of $\hat{D}_i^a(t)$ are always positive (although it is not significantly so for one case) indicates that players hold a stronger belief about their opponent playing a when they just played $\tilde{a}$ than when they chose another action. Our interpretation of this finding is that players think that their opponents try to predict future actions from the past history of the game, and will consequently put more weight on the probability that a will be played again in future rounds, and thus play $\tilde{a}$ more frequently. Turning now to the difference of belief formation between the two treatments, the test-statistics presented in Table 1 are always larger in the Partner than in the Stranger treatment. While the null hypothesis of equality of the distribution means is rejected in all but one case in the Partner treatment, it fails to be rejected in all cases in the Stranger treatment. Overall, these results indicate that when players know they will face the same opponent in the next round, they anticipate a greater propensity for the opponent to play a best response to what has just been played than implied by the history of the opponent's past actions. In other words, they seem to believe that the opponent partly bases his actions on the past history of the game. Because in the Stranger treatment players know that pairs of players are rematched from round to round, they have no reason to put a greater probability on their opponents best responding to their past actions, which implies that the differences between true and empirical beliefs do not differ according to their previous action.

Table 1: Testing for differences in residual beliefs

	Partner treatment			Stranger treatment		
	$\hat{R}_i^X(t)$	$\hat{R}_i^Y(t)$	$\hat{R}_i^Z(t)$	$\hat{R}_i^X(t)$	$\hat{R}_i^Y(t)$	$\hat{R}_i^Z(t)$
$\bar{\hat{D}}_i^a(t)$	0.026	0.062	0.053	0.004	0.013	0.013
t-test	1.380	4.634^{**}	3.723^{**}	0.212	0.963	0.916
	(0.170)	(0.000)	(0.000)	(0.832)	(0.336)	(0.360)
Mann-Whitney	1.977^{*}	5.528^{**}	4.645^{**}	0.282	0.987	1.387
	(0.048)	(0.000)	(0.000)	(0.778)	(0.324)	(0.165)
Somer's D	$0.104^{\dagger}$	0.198^{**}	0.170^{**}	0.018	0.035	0.051
	(0.082)	(0.000)	(0.000)	(0.793)	(0.324)	(0.160)

Note: p-values are shown in parentheses
**, * and † denote significance at the 1, 5 and 10% level

3.2 Teaching a learner

If, as we show in the previous subsection, players are aware of their opponents' learning process, they have an incentive to "teach" adaptive players by choosing actions with

poorer short run payoffs, but which will modify an adaptive opponent's behavior in a way that might lead to higher payoffs in the longer run.

3.2.1 Descriptive analysis

A way of testing if players use a teaching strategy is to examine, along the lines of Ehrblatt et al. (2006), whether their behavior is consistent with their beliefs about their opponents' actions, or if they depart from such a immediate payoff maximizing behavior.

Indeed, the existence of a teaching strategy implies that players will not necessarily best respond to their current beliefs about their opponent's behavior. More precisely, if players anticipate a higher future payoff by playing an action which is not in their immediate best-response set instead of best responding, they might exhibit a stronger tendency to play this action than implied by their current beliefs. We will refer to such a behavior as an 'over response' behavior. From another point of view, one could also say that players 'under respond' by exhibiting a weak tendency to play some actions even if they are in their best-response set. We also expect this over- (and under-) responding behavior to be much weaker in the Stranger than in the Partner treatment. Moreover, because type-1 players have a weaker incentive to use a teaching strategy, their over-response rate should be lower than for type-2 players.

We now attempt to test this implication of the existence of a teaching strategy. For each round, we first derive the belief-wise best-response defined by $argmax_{a\in\{X,Y,Z\}} E_i^a(t)$. Where $E_i^a(t)$ is player i's expected payoff induced by playing action a in round t. For each action X, Y and Z, and for each round, we then calculate the frequency with which it is a best response; as well as the frequency with which the action has actually been played. The difference between these two frequencies gives the rate of over response (which will be negative if players actually under respond). Figures 1, 2 and 3 plot this difference for each possible action, separately for each type of players. They show that the behavior of type-1 players is remarkably similar across treatments, while type-2 players tend to behave differently according to whether they have the opportunity to teach their opponent or not.

In the presence of a teaching strategy, we expect players to have higher rate of over (or under) response in the Partner than in the Stranger treatment because their incentives to deviate from their immediate best response is obviously higher when they repeatedly face the same opponent. According to the results of the previous subsection, the over-response rate should be higher for type-2 than for type-1 players, again because the asymmetry of the payoff matrix implies an asymmetry in their teaching incentives. Note, however, that we do not expect players in the Stranger treatment to play exactly as if they were maximizing their expected immediate payoff in each round. Indeed, other considerations such as fairness or some kind of social norm could induce players to depart from a pure payoff-maximizing behavior. Nonetheless, the incentives brought by the Partner matching protocol should lead to statistically significant differences in the over-response behavior between the two treatments.

Table 2 presents, in its third (for the Partner treatment) and fourth (for the Stranger one) columns, for each possible action and type of players, the mean (over rounds) difference between the frequency with which the action is a best response and the

frequency with which it has actually been played. A positive value indicates that the action has been played more often than implied by players' beliefs in the current round, and a negative value represents the fact that it has been played less often. Significance levels of a test for nullity of these means are displayed next to the over-response rate. The last three columns present test statistics and p-values for various tests (t-test, Mann-Whitney and Somers' D) of equality between the means presented in the two previous columns.

Table 2: Over response

Action	type	Partner	Stranger	t-stat (p-value)	Mann-Whitney (p-value)	Somer's D (p-value)
X	type-1	0.075^{**}	0.096^{**}	-1.231 (0.223)	1.836 (0.066)	-0.275 (0.069)
X	type-2	0.127^{**}	0.071^{**}	3.293 (0.002)	-3.208 (0.001)	-0.481 (0.000)
Y	type-1	0.007	$0.035^{\dagger}$	-1.019 (0.313)	1.086 (0.278)	-0.162 (0.288)
Y	type-2	0.133^{**}	0.053^{**}	3.641 (0.001)	-3.564 (0.000)	0.531 (0.000)
Z	type-1	-0.082^{**}	-0.131^{**}	1.518 (0.134)	-1.764 (0.078)	0.264 (0.078)
Z	type-2	-0.261^{**}	-0.123^{**}	-4.842 (0.000)	4.388 (0.000)	-0.654 (0.000)

Significance levels : †: 10% $*$: 5% $**$: 1%

Although players seem to systematically depart from a best-responding behavior in the Stranger treatment (the means of the over-response rate are almost always significantly different from zero), results show a clear difference in the behaviors of type-1 and type-2 players. While type-1 players seem closer to an immediate expected payoff maximizing behavior in both treatments (their over-response rates are close to zero in both treatments, and are not statistically different between treatments at the 5% level); type-2 players exhibit a behavior that is consistent with the existence of a teaching strategy. While they tend to play X and Y more frequently (and Z less frequently) than implied by their beliefs in the Partner treatment, all these over-response rates become significantly closer to zero in the Stranger treatment. These results are consistent with the fact that type-2 players have stronger teaching incentives than type-1 players.

Figure 1: Over-response rate for action X

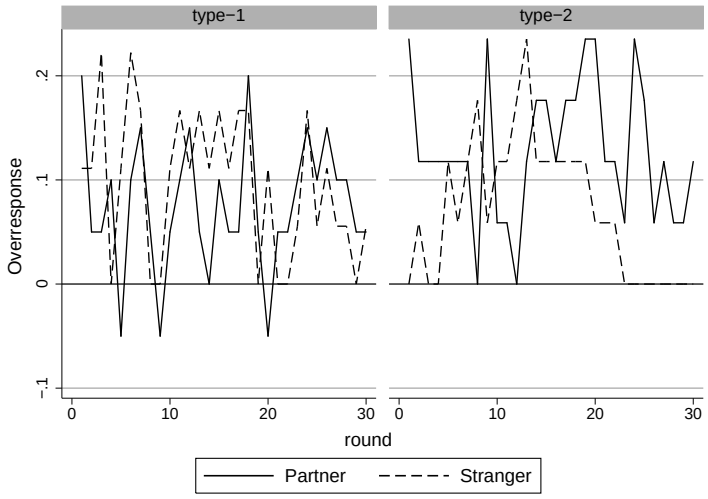


Figure 2: Over-response rate for action Y

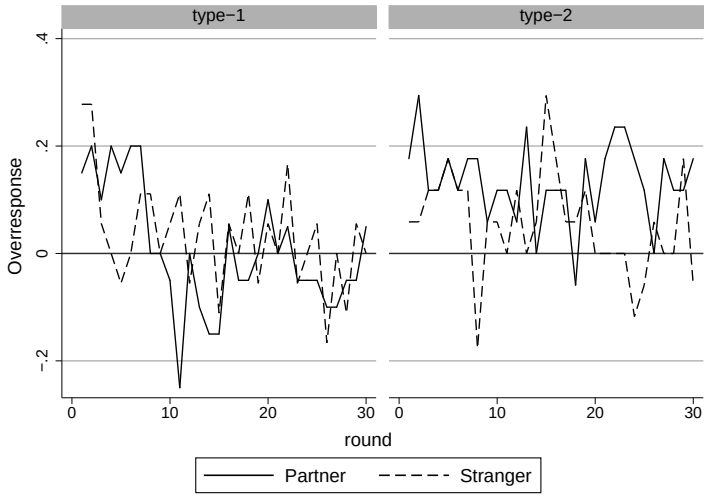
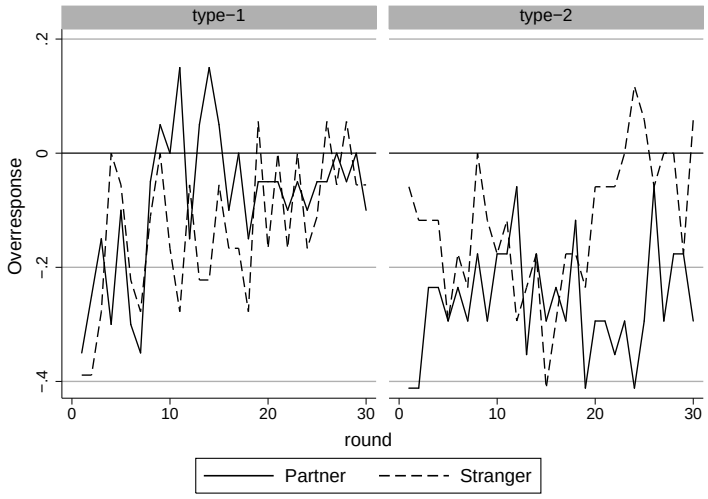


Figure 3: Over-response rate for action Z



3.2.2 Probabilistic choices

Model and estimation In this section, we modify learning models by adding a forward-looking component and see if it helps explaining players' behavior. We first present a rationale on how players might evaluate their gains from using a teaching strategy; and we then turn to the empirical specification and estimation results.

Playing against an adaptive opponent implies that every action played will modify the opponent's future behavior. In order to evaluate the prospects brought by this potential influence, and to decide whether to make use of a teaching strategy, players must first assess the potential gain that would stem from a deviation from their immediate best response.

Playing action k instead of a given reference action r in period t and hence modifying the opponent's behavior in $t+1$ relative to what it would have been if r had been played, will lead to a difference in the expected payoff in round $t+1$. Let δ_i^k denote this perceived difference in player i's expected payoffs when he plays k instead of r. Since an adaptive player's beliefs are updated in every round; the influence of past actions on the opponent's beliefs and behavior will decrease with time. Along the lines of Cheung and Friedman (1997)'s model of varying influence of past actions, we assume that players think this influence will decrease geometrically at rate $\beta \in [0,1)$. If $\beta \to 0$, then the influence of an action played in round t will tend to be only effective for the next round, and will have vanished in subsequent rounds. If $\beta \to 1$, then the potential gains from an action played at t will tend to be carried forward unaltered until the end of the game.

Since players must evaluate the overall gain of playing action k in round t, they have to assess the cumulative expected payoff gain, which can be written as:

$$\theta_i^k = \sum_{\tau=0}^{T-t-1} \delta_i^k \beta^\tau = \delta_i^k \frac{1-\beta^{T-t}}{1-\beta} \tag{2}$$

The way we introduce this forward-looking component is similar to the approach of Rutström and Wilcox (2006). It represents to us an elegant and parsimonious way to incorporate teaching. However, the authors use the same rate to describe a decrease in players' perceived influence on the opponent's beliefs (in their set-up, players assume their opponents are γ-weighted learners who compute their beliefs according to Equation 1) and a corresponding decrease in players' perceived expected payoff gain, represented respectively by γ and β in the present work. We do not follow this intuition here since there is no reason for these two parameters to be necessarily equal.[8] Hence, we introduce this new β-parameter to describe the decreasing rate of the influence of past actions, the interpretation of which is given in the above discussion.

If, as our results of section 3.1 suggest, players think that their opponents modify their beliefs (and their behavior) according to the history of the game; then they have an incentive to use a teaching strategy (i.e. to play an action that does not maximize their expected payoff at the current round, but increases the probability of convergence to an equilibrium they deem preferable). The analysis of Section 3.2.1 has shown that players' behavior was consistent with the existence of such a strategy. We now turn to a more formal test of this hypothesis by fitting a choice model that includes the cumulative expected payoff bonus δ_i^k.

Our empirical model assumes that player i chooses his actions according to (1) the intertemporal expected payoff from playing a given strategy, consisting of (i) the immediate payoff, and (ii) the cumulative expected future payoff difference induced by playing the action, as defined above by θ_i^k; and by (2) an intrinsic attraction for a given action denoted α_i^a to which the player is attracted for non-pecuniary reasons (such as fairness, or other social norms that could influence players' behavior).

Formally, player's i attraction for action a after period t has taken place is denoted $A_i^a(t)$ and can be written:

$$A_i^a(t) = \alpha_i^a + \lambda_i \left[E_i^a(t) + \theta_i^a\right]$$

Where α_i^a is the intrinsic attraction for action a and λ_i represents the player's 'responsiveness' to his expected (immediate and prospective) payoffs.

Some remarks about the way this model formally extends previous belief-based learning models are worth noting. These usual models frequently postulate that players' attraction for an action at a given time depends linearly on the immediate expected payoff induced by this action and on a "bias parameter". This latter parameter is sometimes viewed as reflecting players' non-pecuniary motives or intrinsic attraction for the action. On the other hand, Battalio et al. (2001) argue that this bias might also reflect players' attempts to drive coordination to their favorite equilibrium, which is the kind of behavior we aim to exhibit in this paper. The way we build players' attractions might

[8] Indeed, the way a change in player j's distribution of beliefs will finally impact player i's expected payoff gains is not necessarily that straightforward. Briefly, a change in player j's distribution of beliefs does not necessarily induce the same change in his choice probabilities, which in turn does not necessarily induce the same change in player i's expected payoff gains.

identify the two effects supported by these two interpretations. More precisely, usual attractions take the form $\rho_i^a + \lambda_i E_i^a(t)$; here we use a new specification for the bias parameter ρ_i^a, which allow us to identify teaching: $\rho_i^a = \alpha_i^a + \lambda_i \theta_i^a$. By adding θ_i^a, we might indeed identify the effect suggested by Battalio et al. (2001) and then α_i^a only reflect players' non-pecuniary motives.

Following Fudenberg and Levine (1998) and many others in the literature, we assume that the probability of a given action being chosen in round t takes a logistic form. To identify such models, one has to define a reference action (action X in our application) and normalize its attraction to 1. Parameters of the components of other actions' attraction must thus be interpreted as the difference with the reference action's parameters: α^a is the difference between a's and X's intrinsic attractions; $E_i^a(t)$ is the difference between expected payoffs of a and X given the (true) beliefs.

$$p_i^X(t) = \frac{1}{1 + \sum_{q=Y,Z} \exp\left(A_i^q(t)\right)}$$

$$p_i^Y(t) = \frac{\exp\left(A_i^Y(t)\right)}{1 + \sum_{q=Y,Z} \exp\left(A_i^q(t)\right)}$$

$$p_i^Z(t) = \frac{\exp\left(A_i^Z(t)\right)}{1 + \sum_{q=Y,Z} \exp\left(A_i^q(t)\right)}$$

Where $p_i^X(t)$ is the probability that X will be chosen in round t; $p_i^Y(t)$ and $p_i^Z(t)$ are analogously defined.

As Wilcox (2006) has shown, pooled estimation of such learning models can lead to severe bias in the parameters in the presence of heterogeneity in λ. His Monte-Carlo study suggests that a random-coefficient approach, even if the heterogeneity distribution is misspecified, greatly reduces such bias.

We follow this approach, and use a random-coefficient model, where λ is assumed to follow a two-parameter gamma distribution.[9]

The final log-likelihood is thus:

$$LL = \sum_{i=1}^{N} \sum_{t=1}^{T} \ln \left(\int \left[p_i^X(t)\right]^{\mathbb{1}\{a_i(t)=X\}} \left[p_i^Y(t)\right]^{\mathbb{1}\{a_i(t)=Y\}} \left[p_i^Z(t)\right]^{\mathbb{1}\{a_i(t)=Z\}} f(\lambda)\, \mathrm{d}\lambda \right) \quad (3)$$

where $f(\lambda)$ is the pdf of the Gamma distribution: $f(\lambda; k, \nu) = \lambda^{k-1} \frac{\exp(-\lambda/\nu)}{\Gamma(k)\nu^k}$, and $\Gamma(.)$ is the Gamma function.

The integral in equation (3) is computed using a 32-node Gauss quadrature, and the likelihood is then maximized using standard techniques.

Results We estimate the model of equation (3) separately for the Partner and the Stranger treatments. Since the set-up is not symmetric, we have distinguished between

[9] Which ensures that $\lambda_i > 0 \ \forall i$.

type-1 and type-2 players for the parameters that are likely to differ between types of players, because of the asymmetry in the payoff matrix (intrinsic attractions and prospective payoffs). Other 'psychological' parameters such as β and λ are constrained to be equal across types of players since they should not be influenced by payoffs.

Table 3: Estimation results

Variable	Partner **Coefficient** (Std. Err.)	Stranger **Coefficient** (Std. Err.)
Intrinsic attractions		
α^Y (type-1)	2.564^{**} (0.306)	1.410^{**} (0.238)
α^Y (type-2)	0.218 (0.204)	1.420^{**} (0.415)
α^Z (type-1)	2.648^{**} (0.298)	1.028^{**} (0.223)
α^Z (type-2)	1.131^{**} (0.158)	3.011^{**} (0.378)
Inertia parameter		
β	0.640^{**} (0.034)	0.858^{**} (0.054)
Prospective payoffs		
δ^Y (type-1)	-0.668^{**} (0.101)	0.097 (0.187)
δ^Y (type-2)	3.149^{**} (0.285)	0.643^{*} (0.251)
δ^Z (type-1)	-1.400^{**} (0.147)	$-0.268^{\dagger}$ (0.146)
δ^Z (type-2)	-1.406^{**} (0.171)	-0.833^{**} (0.300)
Heterogeneity parameters		
k	0.147^{**} (0.030)	0.378^{*} (0.150)
θ	11.322^{*} (4.929)	1.000 (0.717)
N	1110	1050
Log-likelihood	-937.055	-797.917
$\chi^2_{(4)}$	148.875	52.581

Significance levels : $\dagger$: 10% $*$: 5% $**$: 1%

Estimated parameters in the Partner treatment (shown in the first column of Table 3) are broadly consistent with the existence of a teaching strategy. Although type-1 players seem to fear that playing the leadership action Y instead of X might lead to a slightly lower future payoff[10], type-2 players expect a significantly positive future payoff gain from playing Y instead of X. Both types of players anticipate a significant future payoff loss from playing the follower action Z. The ranking of expected future payoff differences resulting from playing Y is consistent with the ranking of payoffs in

[10]Possibly to avoid the 'leaders' warfare' (Y,Y) which is most unfavorable for type-1 players.

the underlied equilibria: while type-1 players have little to gain from moving from the (X,X) equilibrium to the (Y,Z) equilibrium, the potential gain is much larger for type-2 players if equilibrium (Z,Y) is reached. However, while the payoff matrix should lead to a lower disincentive for type-2 players to play Z, estimated δ^Z indicates that the expected payoff gains, although negative, are not significantly different between type-1 and type-2 players in the Partner treatment.

Because players are randomly rematched in each round in the Stranger treatment, incentives for teaching should be weaker in this treatment than in the Partner treatment.[11] Hence, we expect the prospective payoff parameters to be much smaller in absolute value in the Stranger treatment than in the Partner treatment. The second column of Table 3 shows that type-1 players in the Stranger treatment do not seem to foresee any significant payoff gain from playing Y instead of X, and a small (and only significant at the 10% level) loss from playing Z instead of X. Type-2 players expect slightly larger differences in future payoffs, but these expected payoff bonuses are much smaller than in the Partner treatment. Note that some form of "social learning" where players learn about the behavior of the population of opponents (as opposed to their individual opponents in the Partner case) might give rise to a corresponding "social teaching" behavior that could explain the small but significant prospective payoff parameters in the Stranger treatment.

While the impact on the opponents' behavior is unsurprisingly thought to be larger in the Partner treatment than in the Stranger treatment, the β-parameters indicate a stronger inertia in the Stranger treatment in comparison with the Partner treatment. Behaviors might indeed evolve faster in a fixed pair of players than in a large population. A reason for this is that players might be quicker in adjusting their beliefs and behavior in a situation where they interact repeatedly with the same opponent compared with a situation where they have to gather information about their current opponent from a population of players.

Finally, the 'intrinsic attraction' parameters show that type-1 players equally favor Y and Z over X, and type-2 players favor Z over the two other actions. This might indicate a preference for equality that we will see reflected in the results of Section 3.3.

3.3 Teaching as a coordination device

The literature on learning models generally aims to track players' behavior without investigating the effective impact of such a behavior on the outcome achieved. Allowing for the use of a teaching strategy might permit to gain some insights on this question. In other words, do players succeed in teaching their opponent to play a given action? More precisely, applying this reasoning to the present set-up, we can ask the following question: Is equilibrium selection affected when teaching is made riskier?

Our results of the previous subsections suggest that type-2 players use this opportunity more than type-1 players do. As a consequence, we expect players to converge

[11] Note that in the Stranger treatment, we had 18 subjects in each session, so the probability of being rematched with the same opponent in the next period is almost 11 per cent, i.e. relatively low. Because of space constraints, we could not run a session with more than 18 subjects and thus we could not decrease this pourcentage.

to type-2 leadership equilibrium (i.e. to the equilibrium that brings type-2 players the highest payoff) more often in the Partner than in the Stranger treatment.

Figures 4 and 5 graph the evolution of the number of equilibria attained in the Stranger and Partner treatment, respectively. Because coordination is easier to achieve in a Partner set-up where players repeatedly interact with the same opponents, differences in the number of equilibria attained might only reflect differences in relative easiness to converge to an equilibrium rather than differences in equilibrium selection due to the existence of a teaching strategy. Thus, in Figures 6 and 7, we rather represent the evolution of the proportion of the various equilibria among the equilibria attained in each round in the Stranger and Partner treatments.

Figure 4: Number of equilibria, Stranger treatment

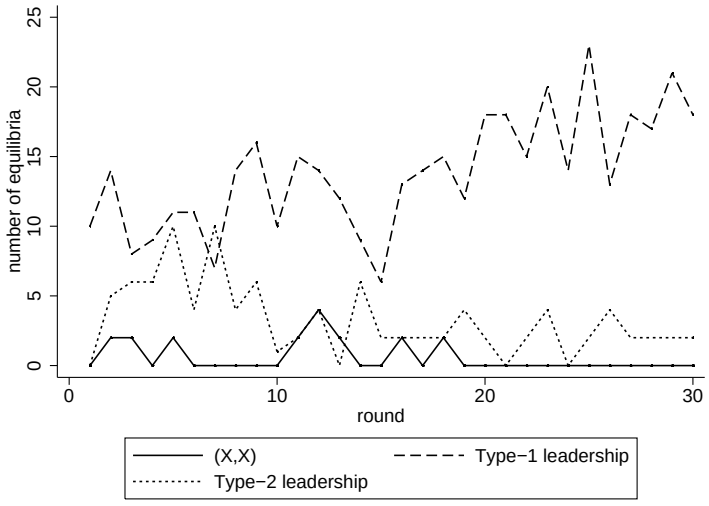


Figure 5: Number of equilibria, Partner treatment

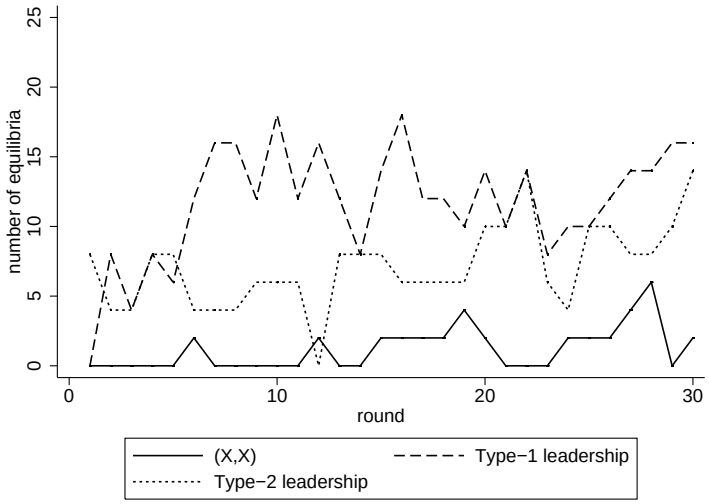


Figure 6: Proportion of equilibria, Stranger treatment

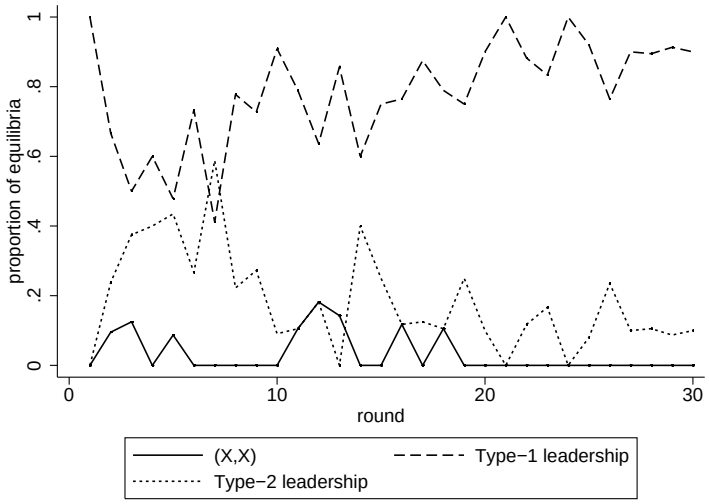
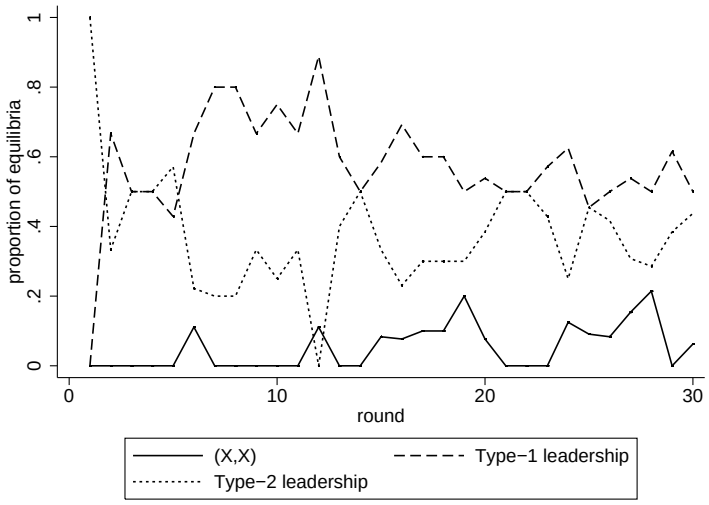


Figure 7: Proportion of equilibria, Partner treatment



Both sets of Figures show that the number and proportion of leadership equilibria where type-2 players are leaders is larger in the Partner than in the Stranger treatment (43.75% in the last round in the Partner treatment, versus 10% in the Stranger treatment), the reverse being true for type-1 leadership (50% in the last round in the Partner case, versus 90% in the Stranger case). Moreover, these graphs show a decrease in the emergence of type-2 leadership equilibrium in the Stranger treatment (from an average of 5.77 equilibria, or 32% during the first 10 rounds, to an average of 2 equilibria, or 9.9% during the last 10 rounds); and an increase in the Partner treatment (from an average of 5.33 equilibria, or 34% during the first 10 rounds, to an average of 9.45 equilibria, or 34.6% during the last 10 rounds[12]); an opposite variation being observed for the emergence of type-1 leadership equilibrium. The (X,X) equilibrium is also more likely to be attained in the Partner treatment, but the difference between treatments is smaller than for leadership equilibria (6.25% versus 0% in the last round).

Table 4 shows the mean count and proportion of each of the 3 equilibria, in both treatments, as well as results from Student, Mann-Whitney and Somer's D tests for the equality of means across treatments.

These results show that the matching protocol does significantly change the distribution of attained equilibria, both in proportion and number. Moreover, these changes are consistent with our previous findings that type-2 players are more prone to use a teaching strategy in comparison with type-1 players. As a result, the number of type-2 leadership equilibria is much higher in the Partner treatment, mainly to the detriment of type-1 leadership equilibria.

[12]The small increase in the proportion of type-2 leadership equilibria in the Partner treatment is partly due to the fact that the only equilibria attained in the first round were 8 type-2 leadership equilibria, thus leading to a proportion of 100% in the first round.

Table 4: Coordination

Equilibrium	Statistic	Partner	Stranger	t-stat (p-value)	Mann-Whitney (p-value)	Somers's D (p-value)
(X,X)	Count	1.2	0.6	10.761 (0.000)	10.210 (0.000)	0.212 (0.000)
(X,X)	Proportion	0.053	0.032	8.254 (0.000)	7.920 (0.000)	0.168 (0.000)
Type-1 leadership	Count	11.733	13.833	−12.183 (0.000)	−10.126 (0.000)	−0.246 (0.000)
Type-1 leadership	Proportion	0.575	0.784	−32.389 (0.000)	−27.099 (0.000)	−0.659 (0.000)
Type-2 leadership	Count	7.133	3.266	33.038 (0.000)	28.544 (0.000)	0.687 (0.000)
Type-2 leadership	Proportion	0.372	0.184	28.567 (0.000)	26.580 (0.000)	0.648 (0.000)

Although type-1 leadership predominates in both treatments (possibly because players have in mind a norm of equality, (Y,Z) being the equilibrium where the differences between players' payoffs is minimal[13]), the opportunity of a teaching strategy in the Partner treatment leads to a doubling (both in absolute value and in proportion) of the number of type-2 leadership. Equilibrium selection is significantly affected by the matching protocol in a way that, as in our previous subsections, is consistent with the use of a teaching strategy by type-2 players, i.e. those who have the greatest incentive to be an efficient teacher, because they have the most to gain from a deviation from a standard immediate payoff-maximizing behavior.

4 Conclusion

Adaptive-learning models have proven to be successful in describing how people behave in games. Yet, in these models, players only look at the past history of the game to choose their current actions, so they do not pay attention to the fact that these current actions could possibly influence their opponents' behavior in the future. In other words, players do not try to outguess their opponents and consequently they do not use a teaching strategy by choosing actions that do not necessarily maximize their immediate expected payoff but might lead to a preferable outcome in the future. Taking this consideration into account might help to track players behavior more accurately and might also allow to gain some insights in predicting the attained outcome of a game.

This paper has used experimental data to examine whether players use a teaching strategy aimed at modifying their opponents' beliefs and actions in order to reach a preferable outcome. We ran a 'Partner' treatment where players were matched in fixed pairs during the whole game, and a 'Stranger' treatment where players were randomly rematched at each period; and had thus no incentive to 'teach' their opponent. Our

[13] As noted in the previous subsection, this social norm, while not formally tested in this paper, is reflected in the 'intrinsic attraction' parameters α^a of Section 3.2.2. These parameters indicate that, regardless of their beliefs, players have a tendency to prefer the (Y,Z) equilibrium, this preference being larger in the Partner treatment than in the Stranger one.

results indicate that players indeed use a teaching strategy and suggest that these considerations are relevant to determine the outcome achieved in a game.

We first tested whether players thought of their opponents as belief-learners who base their beliefs on the past history of the game, which is a necessary condition for players to use a teaching strategy. Our results show that players anticipate their opponents' reaction to their own actions; and take this reaction into account into their own beliefs.

In a second step, we examined whether players actually try to take advantage of this knowledge that they can influence their opponents' beliefs and actions, i.e. we checked whether players' behavior is consistent with the existence of a teaching strategy and whether there is a treatment effect on the tendency of players to use a teaching strategy. We found that, especially when the treatment favors the emergence of teaching, players are likely to depart from a best-responding behavior and choose actions which support a preferable outcome. More particularly, we found that players who have the strongest teaching incentives also have a greater tendency to depart from a best-responding behavior. We then estimated a logistic model which confirmed this tendency. More precisely, the model suggests that when given the opportunity to teach their opponents, the cumulative expected payoff that players could gain by modifying their opponent's behavior had a statistically significant influence on their propensity to play a given action. Again, the model highlights a greater propensity for type-2 players to base their actions on a teaching strategy.

Finally, we investigated the effective relevance of teaching on equilibrium selection and found that teaching indeed drives coordination significantly so that more efficient teachers are more likely to make their favorite equilibrium emerge when they can directly teach their opponents.

References

Arthur, W. B., 1991. Designing economic agents that act like human agents: A behavioral approach to bounded rationality. American Economic Review 81 (2), 353–359.

Battalio, R., Samuelson, L., Van Huyck, J., 2001. Optimization incentives and coordination failure in laboratory stag hunt games. Econometrica 69 (3), 749–764.

Boylan, R. T., El-Gamal, M. A., 1993. Fictitious play: A statistical study of multiple economic experiments. Games and Economic Behavior 5 (2), 205–222.

Camerer, C., Ho, T.-H., 1999. Experience weighted attraction learning in normal-form games. Econometrica 67 (4), 827–874.

Camerer, C., Ho, T.-H., Chong, J.-K., 2002. Sophisticated ewa learning and strategic teaching in repeated games. Journal of Economic Theory (104), 137–188.

Cheung, Y.-W., Friedman, D., 1997. Individual learning in normal form games: Some laboratory results. Games and Economic Behavior 19 (1), 46–76.

Ehrblatt, W. Z., Hyndman, K., Ozbay, E. Y., Schotter, A., 2006. Convergence: An experimental study of teaching and learning in repeated games, mimeo.

Fudenberg, D., Levine, D., 1998. Theories of Learning in Games. MIT Press, Cambridge MA.

Hamilton, J. H., Slutsky, S. M., 1990. Endogenous timing in duopoly games: Stackelberg or cournot equilibria. Games and Economic Behavior 2 (1), 29–46.

Mookherjee, D., Sopher, B., 1994. Learning behavior in an experimental matching pennies game. Games and Economic Behavior 7 (1), 62–91.

Mookherjee, D., Sopher, B., 1997. Learning and decision costs in experimental constant sum games. Games and Economic Behavior 19 (1), 97–132.

Nyarko, Y., Schotter, A., 2002. An experimental study of belief learning using elicited beliefs. Econometrica 70 (3), 971–1006.

Rankin, F. W., Van Huyck, J. B., Battalio, R. C., 2000. Strategic similarity and emergent conventions: Evidence from similar stag hunt games. Games and Economic Behavior 32 (2), 315–337.

Roth, A., Erev, I., 1995. Learning in extensive games: Experimental data and simple dynamic model in the intermediate term. Games and Economic Behavior 8 (1), 164–212.

Rutström, E. E., Wilcox, N. T., 2006. Stated beliefs versus empirical beliefs: A methodological inquiry and experimental test, mimeo.

Sonnemans, J., Offerman, T., 2001. Is the quadratic scoring rule behaviorally incentive compatible?, mimeo.

Wilcox, N. T., 09 2006. Theories of learning in games and heterogeneity bias. Econometrica 74 (5), 1271–1292.

Zeiliger, R., 2000. A presentation of regate, internet based software for experimental economics. `http://www.gate.cnrs.fr/~zeiliger/regate/regateintro.ppt`, Lyon: GATE.